



# Europeana – Core Service Platform

## MILESTONE

### MS27 Research and development plan

<b>Revision</b>	1.0
<b>Date of submission</b>	28 August 2015
<b>Author(s)</b>	Antoine Isaac (EF), Valentine Charles (EF), Nuno Freire (EF), Victor-Jan Vos (EF), Hugo Manguinhas (EF)
<b>Dissemination Level</b>	Public



Funded by  
the Connecting Europe Programme  
of the European Union

## REVISION HISTORY AND STATEMENT OF ORIGINALITY

### Revision History

Revision No.	Date	Author	Organisation	Description
0.1	21/4/2015	Antoine Isaac	EF	Gathering first elements
0.2	17/5/2015	Victor-Jan Vos, Valentine Charles, Nuno Freire, Hugo Manguihas, Antoine Isaac	EF	Adding scenarios from SWOT analysis
0.3	12/8/2015	Antoine Isaac	EF	Re-writing in final structure
1.0	28/8/2015	all authors	EF	Finalizing document, answering review comments from Gregory Markus, David Haskiya, Harry Verwayen, Alastair Dunning

### Statement of originality:

This milestone contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

## Table of Contents

<b>1. Introduction</b>	<b>4</b>
<b>2. R&amp;D mission</b>	<b>4</b>
<b>2.1 R&amp;D mission statement</b>	<b>4</b>
<b>2.2 Brief analysis of Europeana's R&amp;D context</b>	<b>4</b>
<b>3. Our ambitions</b>	<b>6</b>
<b>3.1 Source and share innovative technologies in places where technology is best developed and discussed</b>	<b>6</b>
<b>3.2 Turn overlaps into synergies</b>	<b>7</b>
<b>3.3 Know our stakeholders well and openly welcome new promising collaborations</b>	<b>8</b>
<b>3.4 Fit our activities to a changing technological and organizational landscape</b>	<b>9</b>
<b>4. Conclusion</b>	<b>10</b>

## 1. Introduction

This document presents the strategic goals that will drive Europeana's R&D efforts over the next 5 years. It has been defined through the following process:

- **Identify stakeholders:** we have identified partners and 'customers' of Europeana's R&D and our contribution to them;
- **Analyze the R&D context:**
  - Align with the Europeana 2020 Strategy<sup>1</sup>; especially connect with key priorities (improve data quality, open the data and create value for partners) and core values (usable, mutual and reliable).
  - Identify contributions to the objectives of the Europeana Foundation teams;
  - Identify ongoing and upcoming projects in which Europeana is involved in, and which include R&D activities.
- **SWOT analysis:** we evaluated the strengths, weaknesses, opportunities and threats involved in the context of the R&D team, which led us to identify the most strategic goals and activities for R&D.

## 2. R&D mission

### 2.1 R&D mission statement

*We are a team of experts on technological innovation for aggregation and dissemination of digital cultural heritage data and its associated services. Notably, we spearhead R&D on data exchange, data quality, multilingualism and search.*

*We engage, advise or accompany providers and users of innovative technologies; we assess technology and disseminate knowledge on the best practices developed as part of this process; we facilitate the uptake by Europeana Foundation of the R&D results of the whole network. Our action follows a networking paradigm, where we rather facilitate and coordinate, promoting knowledge exchange and technology uptake.*

### 2.2 Brief analysis of Europeana's R&D context

Our core 'user group' is CH professionals and academics involved in technological innovation on data aggregation and exchange. But the context we operate in can be further refined and extended. It includes the following stakeholders:

- the other teams that constitute the Europeana Foundation.
- the CH organizations that partner EF, especially Europeana's data providers and aggregators.
- providers of innovative technology and technical R&D actors: universities, companies (incl. vendors), non-profit organizations like Wikimedia or RDA, and other DSIs.
- service providing organizations using and re-using digital CH data, such as general research infrastructures, specialized virtual research environments, publishers & bibliographic agencies.
- "peer" aggregators, either non-EU efforts like DPLA<sup>2</sup>, DigitalINZ<sup>3</sup>, OCLC<sup>4</sup>, or European aggregators that have activities parallel to Europeana (BBC).
- funders: European Commission, member states, non-EU organizations such as the Mellon Foundation.

---

<sup>1</sup> <http://strategy2020.europeana.eu/>

<sup>2</sup> <http://dp.la/>

<sup>3</sup> <http://www.digitalnz.org/>

<sup>4</sup> <https://www.oclc.org/>

*We engage, advise or accompany providers and users of innovative technologies; we assess technology and disseminate knowledge on the best practices developed as part of this process; we facilitate the uptake by Europeana Foundation of the R&D results of the whole network. Our action follows a networking paradigm, where we rather facilitate and coordinate, promoting knowledge exchange and technology uptake.*



In this context, doing R&D at the Europeana level is crucial in order to bridge the following gaps.

Europeana and CH organizations can find it hard to benefit from existing R&D for their development.

- Problems for sharing and accessing digital CH are difficult to solve, and often have never been handled.
- The solutions they have to choose need to be state-of-the-art.
- Resources for surveying and assessing technological options in the CH sector are scarce, especially for organizations working at national levels.
- CH stakeholders often do not have a focus on (and lack human resources for) communicating with the tech R&D community and interpreting its results in terms of development opportunities.
- The CH actors with significant technical R&D resources (e.g., BnF, BBC, British Museum, KB...) have to focus on their own organization's requirements.

Reciprocally, R&D technologists often find it hard to characterize the problems of Europeana and CH institutions. We need to spell out cases and defend requirements.

Actors are fragmented, communication and coordination can be improved.

- Many initiatives and projects are taking place across countries and CH sectors.
- Project-based funding and diversity of actors makes it difficult to sustain coherent work.

This creates a need for people to coordinate and create synergies. The Europeana Network is a good base but a coherent and sustained technical innovation will not happen by itself. We must keep track of what has been done and who did it, as well as make time to advise or get involved in relevant initiatives at the moment it is needed.

#### ***How R&D serves Europeana's priorities***

**Priority #1 "Improve Data quality"**. The R&D team has worked on getting richer data in Europeana thanks to an appropriate data model, and our commitment is to continue building and enhancing a 'semantic layer' on top of Europeana objects. Trust in data and content (notably via quality assessment and provenance tracking) will remain high on our agenda.<sup>5</sup>

<sup>5</sup> <http://pro.europeana.eu/page/data-quality-etech15-roundtables>

**Priority #2 "Open the data".** We help design and maintain data models and exchange interfaces (APIs and LOD) and facilitate open access to content (media) in ways technically fit for re-use (e.g., using relevant standards). We work on making them progress toward the status of formal or de facto standards in our community. Finally, a significant part of R&D activities focuses on processes to improve search and multilingual access to culture.

**Priority #3 "Create value for partners".** Our work on data modeling and enrichment contributes to deliver better value to Europeana's data providers, directly or indirectly (i.e., through vendors). We coordinate, facilitate and disseminate the work of the technical innovators in our network through community building activities (EuropeanaTech<sup>6</sup>). For them, we act as hub for knowledge exchange, networking and advice, incl. helping to define R&D agendas and projects relevant to Europeana. For other user groups (Creatives and End Users), we advise on specifications and implementation of the services developed by other EF teams. We maintain a special bond with re-users of data for research (e.g., digital humanists) as targeted by Europeana Research.

#### ***Relation to Europeana's core values***

**Reliable.** We are aware of the state-of-the-art in relevant R&D and/or where to find the required expertise in our network. We are able and available to provide a critical view and advice on developments. Work on more efficient, interoperable solutions benefits the longer-term sustainability of technical work in the Network.

**Usable.** We promote a pragmatic approach to technical innovation. New technologies should stay as much as possible backward-compatible and provide a gradual path to adoption. Standardization must be a matter of encouraging best practices, incl. by documenting and disseminating, rather than enforcing formal, monolithic pieces of technology. At both technical and legal levels we need to encourage openness<sup>7</sup>.

**Mutual.** Embedded in the Europeana Network and similar communities, we believe in engaging and coordinating, as well as in the power of multilateral exchange of expertise. Development of usable, innovative technology is not a top-down process and needs multiple iterations. We are experts but we are inclusive and able to communicate at various levels, including producing training and educational material.

### **3. Our ambitions**

#### **3.1 Source and share innovative technologies in places where technology is best developed and discussed**

*We engage our partners in various spaces:*

- projects the Europeana Foundation participates in, as partner or subcontractor
- the Europeana Network, especially the EuropeanaTech community in which we share information and can find expertise to guide Europeana's efforts (e.g., through task forces).
- other projects where R&D team members may e.g. be invited to participate as advisors
- standardization efforts like W3C or DCMI
- professional and academic conferences and publications
- ad-hoc collaborations (e.g., BBC, Latvian ministry of culture<sup>8</sup>)

<sup>6</sup> <http://pro.europeana.eu/get-involved/europeana-tech>

<sup>7</sup> <http://pro.europeana.eu/page/issue-3-open-data-and-data-re-use>

<sup>8</sup> <http://pro.europeana.eu/blogpost/latvian-ministry-of-culture-and-europeana-sign-memorandum-of-und>

### **Improvement of services in Europeana and related projects**

We help the Europeana teams to implement and run an efficient, state-of-the-art portfolio of services, focusing on data exchange and Linked Data, search and multilingual functions, data management, enrichment and re-use.

We do not coordinate the development of products ourselves. We rather assist the definition and evaluation of new products, prototyping, support the Data and Partner Services team for operating our 'most R&D-advanced' services. We especially determine which models, industry and groups should Europeana and its partners be interoperable with. We can play a key role in making sure the solutions developed and adopted in the CH sector stay interoperable and efficient. We also aim at reducing the need for (duplicated) development of technologies within Europeana by showcasing and encouraging the exploitation of solutions favored by the Network, as well as infrastructural services developed in other areas, such as translation, data mining or annotation. We hope this will contribute to better sustainability of technical developments in our community.

*For example, the expertise of the team enables us to:*

- *develop EDM as a framework and keeping 'modeling divergences' in check among various Europeana-related services<sup>9</sup>.*
- *define a 'data strategy' where relevant data resources are exploited to enrich the metadata gathered by Europeana and its partners.*
- *help fitting recommendations from the R&D community into evolving products that support new flows for content aggregation.*
- *explore the relation between data enrichment, Linked Open Data and the new Cloud developments in Europeana Cloud.*
- *investigate using the CEF.AT translation services.*
- *interact with research infrastructures, like CLARIN, to optimize the usage by third parties of Europeana data, to benefit from existing text processing platforms instead of developing one specific to Europeana Cloud.*

### **Community management and knowledge exchange**

Our aim is to help R&D actors from the Europeana ecosystem to identify and communicate better their problems and their solutions, with a practitioner's focus.

The R&D team acts here as matchmaker and incubator for R&D/innovation in our community, for example, by participating and organizing R&D events, and by disseminating relevant research results across communities.

*For example, we organize the EuropeanaTech conference, publish R&D results through Europeana communication channels (Europeana Labs, Europeana Pro, & EuropeanaTech Insight) or at external venues (professional and scientific conferences and journals)*

## **3.2 Turn overlaps into synergies**

*Do we have competitors? We are in a unique position, embedded in the Europeana Network. It is difficult, if not impossible, for others to fulfil the same missions. Other teams involved in similar coordination efforts have approaches that complement ours, rather than overlap with it. Either they serve a different market (OCLC, DPLA) or they play a different (cross-domain) role in the R&D chain (W3C, DCMII). Academic or industrial R&D actors (various universities, R&D departments in large CH orgs, companies like Ontotext) may have similar foci, but they fill different niches in the ecosystem. We actually aim to facilitate their work, help them reach out to our entire network, and concretely collaborate with them when appropriate<sup>10</sup>. Sometimes*

<sup>9</sup> <http://pro.europeana.eu/publication/enhancing-the-europeana-data-model-edm>

<sup>10</sup> See for example our R&D experiment with OCLC Research, <http://pro.europeana.eu/blogpost/hierarchical-clustering-making-sense-of-europeana-data>

*Europeana may work on similar products as developed elsewhere: our role is then to align goals and encourage synergies. Resources in the CH domain are scarce and all actors should seek to provide the best value for their funders.*

### **Community management and knowledge exchange**

We enhance synergies within the Europeana Network, for example by advising Europeana-related projects on technical innovation. We are able to use our skills in steering activities and working groups to make sure the CH sector adopts common, state-of-the-art solutions. We keep a finger on the pulse of R&D activities within the Europeana Network. We are the hub connecting the CH R&D ecosystem to other, possibly more technological communities (e.g., Semantic Web). In particular, we serve as a main reference point for R&D/innovation partners from other sectors (DSIs, DARIAH, RDA and other actors mentioned in CEF), who cannot handle interacting with many individual Europeana Network members. For them we can represent the Network, acting as matchmaker or expertise center. In practice we can do this by collecting use cases and requirements that we present on behalf of the Network.

*For example, the team has the expertise to establish a solid position for Europeana as the European competence center for CH within the research infrastructure landscape of Europe. We are active in their groups and can invite them to relevant Europeana activities.*

Note that we do not need deep knowledge in all areas we are covering. But we must be in position to source the relevant expertise from elsewhere, on request.

### **Steering of research agenda in CH**

We ensure Europeana has a voice in the R&D community. We advise them, and help identifying and solving problems that are key for us.

We try to steer and validate R&D work in projects and other circles (workshops, conferences...) to ensure a maximal level of quality and consistency with respect to Europeana's strategy.

We also help finding new resources, leveraging our community activities to submit new, better project proposals with relevant activities, especially on data management, standardization and best practices. We will use our ability to build communities so as to more pro-actively engage with project proposals or even calls, making sure that they consider our strategy and will work for our purposes.

We must act from an authority position. Our key position in the network must be backed by a culture of technical scrutiny that may not exist in other Europeana and project teams, as well as technical expertise and a broad knowledge of the state-of-the-art and ongoing activities in the Network. To this end, we seek to maintain technology inventories or registries, such as the our inventory of Free, Libre and Open Source Software (FLOSS).

*For example, for the 'semantic technology' area, we foresee especially work on:*

- *metadata mappings and alignments*
- *vocabularies defined and maintained by Europeana, or used in particular projects*
- *vocabularies which Europeana is able to exploit in its data process*
- *vocabulary alignments*

### **3.3 Know our stakeholders well and openly welcome new promising collaborations**

*Our key stakeholders are undoubtedly our colleagues at the Europeana foundation, in the CH organizations or the CH technology providers in the Europeana Network.*

*However this network is open to include any new relevant members, and we shall aim at a healthy growth of our technology-focused community.*

*Furthermore, not all relevant partners strictly belong to the CH domain (e.g., ERICs, projects like FREME, associations like LT-Innovate). Having a dedicated Europeana R&D team is a unique opportunity to project the CH domain beyond its traditional boundaries and benefit from*



*technological advances in other communities. The team provides one single, Europe-level contact point for other communities who would like to address CH problems or re-use CH data.*

As mentioned, our expertise and collaboration with groups working on data quality, normalization and enrichment will contribute to fulfill the DSI objectives.

For the aggregators and Europeana expert hubs, we are in a position of facilitators ('hub of hubs'). Our expertise is not domain-specific (e.g., we are not archaeologists). But we can relate the domains with 'transversal innovation' that benefits them most.

*For example, the transition to the Cloud provides opportunities for new developments for Europeana's infrastructure and EDM. This area is new to the team and relevant collaborations areas might still be unknown. Exploiting the research funding opportunities of H2020 in Europe-wide research infrastructures, would allow Europeana to strengthen its partnerships and achieve greater efficiency by working with solid research infrastructures.*

The team can use its communication strength to reach out new groups. Its positioning in Europe is an asset to get the credibility required for new collaborations.

Our international network should be the opportunity to diversify funding sources for CH innovation that matches Europeana's strategy.

*Oversea collaboration (e.g. with DPLA and other US initiatives) offers great opportunities, e.g., for common platforms and models for the management and exchange of data.*

Our activities should remain funded under a European remit, however. Benefiting from national or international resources for relevant partners' activities can be great, but we must not become biased towards such initiatives.

Across the Network, we must also ensure that the balance of stakeholders in the relevant groups is maintained. This requires very good knowledge of our partners, their resources and their priorities.

### **3.4 Fit our activities to a changing technological and organizational landscape**

*In such a wide and varied group of stakeholders, it is difficult to anticipate where most of the work should be done, at any given time. The evolution of technology and the constant organizational change requires agility. Having a dedicated R&D team with enough flexibility creates the opportunity to adapt to shifting operational priorities and external opportunities and even be proactive.*

#### **Support and advice on collaborations**

We seek to provide forward coordination, but we also must be flexible and seize important collaboration opportunities whenever they occur, or assist other Europeana teams or projects who would like to embark on such opportunities.

In particular, we must balance long-term R&D priorities with short-term development deadlines. Europeana as a Network aims to be at the forefront as a general network, but not all partners are on the same page. Some partners are frontrunners, and can be connected with the relevant communities. Other will need more exposure to the work of such frontrunners.

For example, ongoing work by Ontotext and Europeana on exploiting Wikidata can be seen as a trailblazing effort for connecting other Cultural Heritage organizations to this new collaborative knowledge base.

To play a role in porting relevant R&D results to the 'market', our knowledge transfer and coordination activities have to be swift.

Shifting priorities are unavoidable. But we can make them fit into an orderly strategy, by picking the appropriate level of success for our R&D coordination work for each situation.

*For example, overseas collaboration offers great opportunities, however it brings our work to a scale that might be hard to handle. Such opportunities can in fact aggravate our current weakness in terms of resources and impact development and communication with our own*

*network. But there are several degrees of collaboration possible, from common prototyping activities to advising and knowledge exchange. Or even by letting other teams and partners interacting with the given partner, once the 'R&D stage' has been passed.*

Another answer, which is fit to the level of an R&D team, is to being involved early in the process of collaboration.

We should also seize opportunities for external collaboration in a way that is visible for the organization. Prioritization for external work should always be balanced according to priorities/visibility of that collaboration for another internal stakeholder.

*For example, Wikidata activities should always be tied to (1) data enrichment for search, portal or building channels (2) Europeana's strategic collaboration with the Wikimedia Foundation and community.*

Constant prioritisation should help us pick the right level of engagement and put us in a better position to follow the relevant/key areas of technical innovation. If we focus on the things we're best at and that are the most relevant for our network (starting with the aforementioned EDM, data enrichment, multilinguality), and we can select the best partners to work on them, we will be able to deliver more values to our stakeholders.

Doing this would also put us in the situation of helping in the wider context than R&D alone, by creating mission statements (or briefs) for partnerships, and selecting the right partners.

What makes a 'good' partner on our side depends on the nature of the project. One dimension is the stability of the partner (people working in a group, people leaving). As it has more flexible means of engaging partners, an R&D team will often have had a lot of experience with potential partners, and can advise the identification of partners and the selection of the projects.

## **4. Conclusion**

In line with our emphasis on flexibility and forward looking, this strategy will be maintained as a living document, where we add input from others in the Europeana Foundation and the Network. In the coming months work will be done to implement the strategy into more concrete, thematic plans. We will follow the axes represented by the coming milestones for the R&D stream in the Europeana Digital Service Infrastructure:

*Overall plan to coordinate, innovate and disseminate Europeana coordinated R&D.*

We will trace the history of R&D coordination and dissemination activities (EuropeanaTech) and provide recommendations for ensuring they can operate smoothly and efficiently. We will notably discuss the FLOSS inventory, our recently launched EuropeanaTech Insight publication and, at a more general level, the relations between our EuropeanaTech community activities and general Europeana efforts on dissemination and knowledge sharing (Europeana Pro, Europeana Labs). We will explore ways to participate in R&D project bids and help ensuring that technical developments resulting from R&D are sustainable after project funding ends.

*EDM development plan*

EDM is maintained as a 'living standard' for the Europeana community. New projects and applications sometimes require adapting existing modeling features, or the addition of new extensions. The EDM development plan will sum up the ongoing and planned work. A specific focus will be put in the coming year on exchange annotations on Europeana objects across different platforms, and identifying relevant standard and guidelines for aggregating and distributing media files (e.g., IIIF).

*Search improvement plan*

We will explore the various options to enhance the performance of Europeana's search service, including actions on enriching data with relevant contextual sources, better ranking, but also preparing the ground for a more systematic approach to evaluating the performance of our services. This document, a collaboration between the product development, the technical and the R&D teams, will discuss short-term developments for DSI as well as longer-term R&D plans.

*Multilingual research and development plan*

This plan will expand on the previous ones, bringing in one same effort everything we do to help the Europeana Network perform better services in a multilingual environment. Specific multilingual aspects of data quality and enrichment will be tackled, as well as providing better multilingual search and navigation functions.

**Acknowledgments**

Gregory Markus, David Haskiya, Harry Verwayen, Alastair Dunning have reviewed and made useful suggestions to earlier versions of this plan.